

# ON COMPRESSION USING THE DISTRIBUTED KARHUNEN-LOÈVE TRANSFORM

Michael Gastpar, Pier Luigi Dragotti, and Martin Vetterli

Department of EECS, University of California, Berkeley, CA 94720-1770

Department of Electrical and Electronic Engineering, Imperial College, London, SW7 2BT, UK

Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland

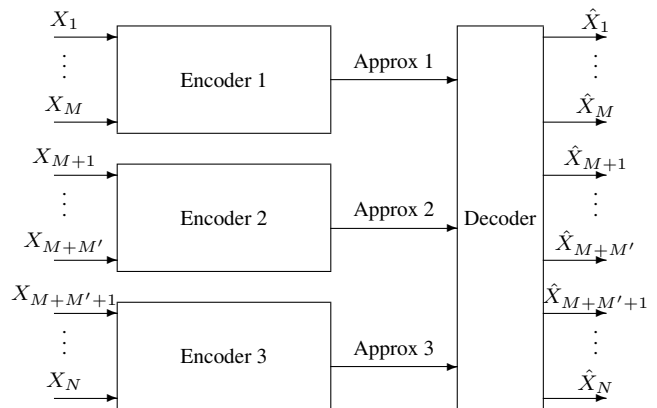
[gastpar@eecs.berkeley.edu](mailto:gastpar@eecs.berkeley.edu), [p.dragotti@ic.ac.uk](mailto:p.dragotti@ic.ac.uk), [martin.vetterli@epfl.ch](mailto:martin.vetterli@epfl.ch)

## ABSTRACT

In this paper, we discuss a framework for the distributed compression of vector sources, based on our previous work on distributed transform coding [1, 2, 3, 4]. In particular, our goal is to develop a strategy of first applying a suitable distributed Karhunen-Loève transform, whereafter each component can be handled by standard distributed compression techniques. In the present paper, we first study the scenario where all but one terminal furnish a noisy approximation of their observation. For the case where the underlying vector is Gaussian, and the added noise is also Gaussian, we establish that indeed, it is optimal for the last terminal to apply a (local) transform to its observations, and to separately compress each component in the transform domain. Then, we outline how this leads to a general simple distributed compression strategy for Gaussian vector sources: Each terminal applies a suitable local transform to its observations, and encodes the resulting components separately in a Wyner-Ziv fashion, i.e., treating the compressed descriptions of all other terminals as side information available to the decoder. This achieves the best known performance. The optimum performance is unknown to date.

## 1. INTRODUCTION

The implementation of distributed compression strategies has recently received considerable attention [5, 6, 7, 8, 9, 10]. In this paper, we address the situation where multiple terminals each sample a part of a large object, characterized by a vector  $X$  of length  $N$ , and the goal is to reconstruct the entire large object to within the smallest mean-squared error possible. For the scope of the present paper, we restrict attention to the case where  $X$  is a vector of jointly Gaussian random variables. One task is to determine the smallest rates required to achieve a fixed distortion level. However, this problem is unsolved to date, even for the case  $N = 2$ . Bounds appear in [11, 12]. Our goal is to extend the currently known best achievable rates to the scenario where each terminal observes a sub-vector of the overall vector  $X$ .



**Fig. 1.** The distributed KLT problem: Distributed compression of multiple correlated vector sources.

We discuss a scheme where each terminal applies a suitable transform, whereafter each component can be compressed separately. Hence, there are two key parts to the solution: the optimal transform, and the bit allocation between the components. These two tasks have to be addressed jointly.

In the present paper, we first study in detail a simplified scenario of the general distributed compression problem, where all but one of the terminals furnish noisy approximations of their observations to the decoder, and the last terminal is asked to optimally compress its observations, considering that the decoder has access to side information. This is a slight variation on the problem studied by Wyner and Ziv [13, 14], the variation being in the *distortion criterion*: the ultimate goal of the overall coding effort is to enable the decoder to furnish the best approximation of the entire vector  $X$ . Then, we outline how this leads to a certain distributed compression strategy. It cannot be shown that this strategy is optimal since the optimum is unknown to date, but it can be shown that this strategy achieves the currently known best performance for the case of a Gaussian random vector  $X$  and a mean-squared error fidelity criterion.

## 2. PROBLEM STATEMENT

The distributed compression problem addressed in this paper is illustrated in Figure 1. There are  $L$  terminals (in the figure,  $L = 3$ ), each observation a part of the vector

$$X \stackrel{\text{def}}{=} X_1, X_2, \dots, X_M. \quad (1)$$

Throughout this paper, we assume that  $X$  is a vector of  $N$  jointly Gaussian random variables, with mean zero and covariance matrix  $\Sigma$ . The assumption that  $X$  has zero mean is not crucial, but it considerably simplifies the notation. Therefore, it is kept throughout the paper. Moreover, throughout the paper, we assume that the fidelity (or distortion) criterion is the mean-squared error between the vector  $X$  and the best estimate  $\hat{X}$  that the decoder can produce based on the approximation furnished by the encoders,

$$D = E\|X - \hat{X}\|^2. \quad (2)$$

## 3. THE CONDITIONAL KLT WITH NOISY SIDE INFORMATION

In this section, we extend our previous results [1, 2, 4] to the scenario illustrated in Figure 2: The decoder has access to side information about the vector  $X$ , as follows. The encoder samples the first  $M$  components of the vector  $X$ , denoted by

$$X_S = (X_1, \dots, X_M), \quad (3)$$

with zero mean and covariance matrix denoted by  $\Sigma_S$ . The side information at the decoder is derived from the remaining  $N - M$  components of the vector  $X$ , denoted by

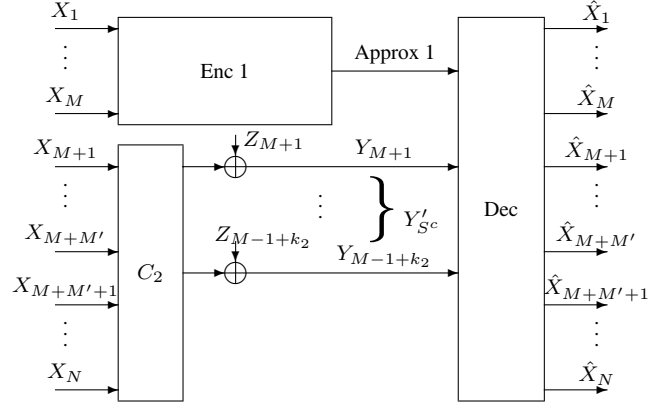
$$X_{S^c} = (X_{M+1}, \dots, X_N). \quad (4)$$

We denote the covariance matrix between  $X_S$  and  $X_{S^c}$  by  $\Sigma_{SS^c}$ . However, the decoder cannot directly access  $X_{S^c}$ . Instead, a transformed and noisy version is available,

$$Y'_{S^c} = C_{2a}X_{S^c} + Z, \quad (5)$$

where  $Z = (Z_{M+1}, \dots, Z_{M-1+k_2})$  denotes the vector of added Gaussian noise. We assume it to be of zero mean and covariance matrix  $\Sigma_Z$ . For most of our derivations in the sequel,  $Z$  need not be independent of  $X_{S^c}$ ; denote their cross-correlation matrix by  $\Sigma_{ZS^c}$ .

**Remark 1.** The additive noises  $Z_{M+1}, \dots, Z_{M-1+k_2}$  can be thought of as modeling the effect of compressing the side information  $X_{S^c}$ . Note that at this point, we are assuming the noise components  $Z_{M+1}, \dots, Z_{M-1+k_2}$  neither to be mutually independent, nor to be independent of the signal  $X_{S^c}$ . However, clearly, the Gaussian assumption does impose certain limitations to our model.



**Fig. 2.** Compression involving partial observation (or sub-sampling) and side information.

As a consequence of the assumption that  $X$  is a vector of jointly Gaussian random variables, and that the additive noise  $Z$  is also Gaussian, we can write

$$X_{S^c} = AX_S + A_2Y'_{S^c} + V, \quad (6)$$

where  $(X_S, Y'_{S^c})$  and  $V \in \mathbb{C}^{N-M}$  are independent Gaussian random vectors, and  $A \in \mathbb{C}^{N-M \times N-M}$  and  $A_2 \in \mathbb{C}^{N-M \times k_2}$  are constant matrices. By the same token, we can write

$$\begin{pmatrix} I_M \\ A \end{pmatrix} X_S = BY'_{S^c} + W, \quad (7)$$

where  $Y'_{S^c}$  and  $W \in \mathbb{C}^N$  are independent Gaussian random vectors, and  $B \in \mathbb{C}^{N \times k_2}$  is a constant matrix.

The goal of our consideration is to minimize the distortion

$$E \left[ \|X - \hat{X}\|^2 | Y'_{S^c} = y'_{S^c} \right]. \quad (8)$$

As we will see, at least in the Gaussian case considered in this paper, the minimizing solution will not require the encoder to know the value of the realization  $y'_{S^c}$ . In line with our previous work [1, 2, 4], two perspectives are of interest to our further considerations: First, the problem of finding the best  $k$ -dimensional approximation, and second, the problem of compressing  $X_S$  using the least number of bits.

### 3.1. Best $k$ -dimensional Approximation

For the problem illustrated in Figure 2, suppose that Encoder 1 may furnish a  $k$ -dimensional approximation of its  $M$ -dimensional observation. The  $k$  dimensions should be picked such as to minimize the distortion criterion (8). What is the best set of  $k$  dimensions? To determine this, we

rewrite the distortion criterion as follows:

$$\begin{aligned} E \left[ \|X - \hat{X}\|^2 | Y'_{S^c} = y'_{S^c} \right] \\ = E \left[ \|X_S - \hat{X}_S\|^2 | Y'_{S^c} = y'_{S^c} \right] \\ + E \left[ \|X_{S^c} - \hat{X}_{S^c}\|^2 | Y'_{S^c} = y'_{S^c} \right] \end{aligned}$$

The key step is to relate the estimate  $\hat{X}_{S^c}$  to the estimate  $\hat{X}_S$ . Since  $\hat{X}_{S^c}$  is the (conditional) minimum mean-squared error estimate of  $X_{S^c}$ , it can be expressed as

$$\hat{X}_{S^c} = E[X_{S^c} | T, Y'_{S^c} = y'_{S^c}],$$

where  $T$  is the approximation furnished by Encoder 1. Using (6), this can be rewritten as

$$\begin{aligned} \hat{X}_{S^c} &= E[AX_S + A_2 Y'_{S^c} + V | T, Y'_{S^c} = y'_{S^c}] \\ &= E[AX_S + V | T, Y'_{S^c} = y'_{S^c}]. \end{aligned}$$

Finally, since  $T$  is only a function of  $X_S$ , it is independent of  $V$ , hence,

$$\begin{aligned} \hat{X}_{S^c} &= AE[X_S | T, Y'_{S^c} = y'_{S^c}] \\ &= A\hat{X}_S. \end{aligned}$$

Using this insight, we find

$$\begin{aligned} E \left[ \|X - \hat{X}\|^2 | Y'_{S^c} = y'_{S^c} \right] \\ = E \left[ \|X_S - \hat{X}_S\|^2 | Y'_{S^c} = y'_{S^c} \right] \\ + E \left[ \|AX_S + V - A\hat{X}_S\|^2 | Y'_{S^c} = y'_{S^c} \right] \quad (9) \end{aligned}$$

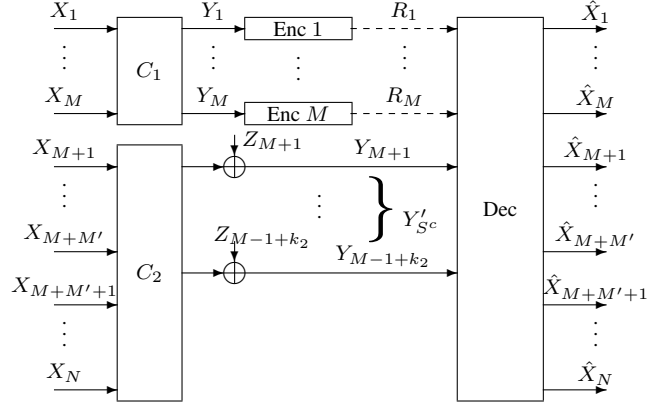
Noting that  $V$  is, by construction, independent of  $Y'_{S^c}$ , and collecting terms, the mean-squared error can be expressed as

$$E \left[ \|(I_A^M)X_S - (I_A^M)\hat{X}_S\|^2 | Y'_{S^c} = y'_{S^c} \right] + E\|V\|^2. \quad (10)$$

The last step is to apply a transform  $C_1$  to  $(I_A^M)X_S$  such that the resulting components, denoted by  $Y_S$ , are conditionally uncorrelated given  $Y'_{S^c}$ . In the considered Gaussian scenario, such a transform exists: it is simply the eigendecomposition of the conditional covariance matrix of  $(I_A^M)X_S$  given  $Y'_{S^c}$ . This conditional covariance matrix is simply the covariance matrix  $\Sigma_W$  of  $W$  as defined in Equation (7). Denote the eigendecomposition of  $\Sigma_W$  by  $Q\Lambda Q^H$ , where  $Q$  is a unitary matrix, and  $\Lambda$  is diagonal. Then,  $Y_S = QX_S$ , and in terms of  $Y_S$ , the distortion can be written as

$$\begin{aligned} E \left[ \|Y_S - \hat{Y}_S\|^2 | Y'_{S^c} = y'_{S^c} \right] + E\|V\|^2 \\ = \sum_{m=1}^M E \left[ |Y_m - \hat{Y}_m|^2 | Y'_{S^c} = y'_{S^c} \right] + E\|V\|^2 \end{aligned}$$

The last expression permits a simple characterization of the solution to the considered problem, as follows.



**Fig. 3.** Compression involving subsampling and side information: An optimal strategy.

**Theorem 1.** *The best  $k$ -dimensional approximation that Encoder 1 can furnish is a subspace of the eigenvectors of the matrix  $\Sigma_W$ . The best  $k$  dimensions are the eigenvectors corresponding to the  $k$  largest eigenvalues  $\lambda_i^2$  of the matrix  $\Sigma_W$ .*

### 3.2. Compression Problem

Reconsider the problem illustrated in Figure 2, but suppose now that Encoder 1 may furnish a description using  $R$  bits per sample (i.e.,  $R$  bits for each observed  $M$ -dimensional vector). Again, this description should be shaped in such a way as to minimize the distortion criterion (8).

**Theorem 2.** *The rate-distortion function for the problem with side information, illustrated in Figure 2, where  $X$  is a vector of jointly Gaussian random variables, is given by*

$$R(D) = \min_{D_i} \sum_{i=1}^M \max \left\{ \frac{1}{2} \log_2 \frac{\nu_i^2}{D_i}, 0 \right\} \quad (11)$$

where  $\nu_i^2$  are the eigenvalues of the matrix  $\Sigma_W$ , and where the minimum is over all  $D_i$  satisfying  $\sum_{i=1}^M D_i + E\|V\|^2 \leq D$ .

## 4. DISTRIBUTED COMPRESSION STRATEGIES

In an information-theoretic setting, the optimal codebooks to achieve the promises of Theorem 2 are Gaussian. The fact that each component in the transform domain is encoded separately means that the encoding operation can be thought of as adding independent (across the  $M$  components) and white (across time) Gaussian noise to the components  $Y_1, Y_2, \dots, Y_M$ . The components that get zero rate in the rate allocation problem (11) are simply omitted.

This insight suggests a simple algorithm to design coding strategies: Once the  $M$  encoding terminals in Figure 3 have been replaced by additive (white) Gaussian noise (or simply omitted whenever  $R_i = 0$ ), the overall situation becomes symmetric, and we can proceed in a round robin fashion: We fix the transform  $C_1$  and the additive noises for Encoder 1, and optimize the transform and the rate allocation for Encoder 2. The convergence of this procedure will be studied later. Notice, however, that the *total* rates for each encoder, denoted by  $R_{(l)}$ , for  $l = 1, 2, \dots, L$ , (i.e., for example, from Figure 3, we see  $R_{(1)} = R_1 + R_2 + \dots + R_M$ ), are *coupled*. In particular, they have to satisfy sum rate conditions that can be derived in extension of the arguments presented in [11]. This compression strategy is analyzed in more detail in [4].

## 5. CONCLUSIONS AND EXTENSIONS

This paper develops the key building block of a distributed compression strategy for Gaussian vector sources. The strategy consists in applying a local transform (a distributed KLT) to each vector source independently, whereafter the components of that vector source are encoded using standard side information encoding. The full strategy will be described and analyzed in [4].

### A. FORMULAE

The matrices  $A$  and  $A_2$  in Equation (6) can be calculated from standard results about multivariate Gaussian distributions. In the simple case where the additive noise  $Z$  is independent of the signal vector  $X$ , this can be expressed as

$$\begin{pmatrix} A \\ A_2 \end{pmatrix} = Cov(X_{S^c}, X) \begin{pmatrix} I_M & 0 \\ 0 & C_{2a} \end{pmatrix}^H \cdot \left( \begin{pmatrix} I_M & 0 \\ 0 & C_{2a} \end{pmatrix} \Sigma \begin{pmatrix} I_M & 0 \\ 0 & C_{2a} \end{pmatrix}^H + \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_Z \end{pmatrix} \right)^{-1}. \quad (12)$$

Similarly, the covariance matrix  $\Sigma_W$  of the random vector  $W$  in Equation (7) also follows from standard results about multivariate Gaussian distributions. In the simple case where the additive noise  $Z$  is independent of the signal vector  $X$ , it can be written as

$$\begin{aligned} \Sigma_W &= \begin{pmatrix} I_M \\ A \end{pmatrix} \Sigma_S (I_M A^H) \\ &\quad - \begin{pmatrix} I_M \\ A \end{pmatrix} \Sigma_{SS^c} C_{2a}^H (C_{2a} \Sigma_{S^c} C_{2a}^H + \Sigma_Z)^{-1} \\ &\quad C_{2a} \Sigma_{SS^c}^H (I_M A^H). \end{aligned} \quad (13)$$

### B. REFERENCES

- [1] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," in *Proc 2002 International Workshop on Multimedia Signal Processing*, St. Thomas, US Virgin Islands, December 2002.
- [2] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loève transforms," in *Proc 2003 IEEE Data Compression Conference (DCC)*, Snowbird, UT, March 2003.
- [3] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *EPFL I&C Technical Report IC/2003/12*, March 2003, To be submitted to *IEEE Trans Info Theory*, Nov. 2003.
- [4] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *to be submitted to IEEE Transactions on Information Theory*, 2003.
- [5] S.S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Data Compression Conference*, Snowbird, Utah, March 1999, pp. 158–167.
- [6] S.S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor network," in *Data Compression Conference*, Snowbird, Utah, March 2000, pp. 363–372.
- [7] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 51–60, March 2002.
- [8] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communication Letters*, vol. 5, pp. 417–419, October 2001.
- [9] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Data Compression Conference*, Snowbird, Utah, March 2002, pp. 252–261.
- [10] A. Liveris, Z. Xiong, and C. Georgiades, "A distributed source coding technique for correlated images using Turbo codes," *IEEE Communications Letters*, vol. 6, pp. 440–442, October 2002.
- [11] T. Berger, "Multiterminal source coding," *Lectures presented at CISM Summer School on the Information Theory Approach to Communications*, July 1977.
- [12] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. IT-43, no. 6, pp. 1912–1923, November 1997.
- [13] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-22, pp. 1–11, January 1976.
- [14] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and Control*, vol. 38, pp. 60–80, 1978.